Appendix 20. Model Archival Summary for Alkalinity Concentration at Station 06892350; Kansas River at De Soto, Kansas

This model archival summary summarizes the alkalinity concentration (Alk) model developed to compute 15-minute Alk from July 19, 2012 onward. This model supersedes all previous models.

Site and Model Information

Site number: 06892350
Site name: Kansas River at De Soto, Kansas
Location: Lat 38°59'00", long 94°57'52" referenced to North American Datum of 1927, in NE 1/4 SE 1/4 SE 1/4 sec.28, T.12 S., R.22 E., Leavenworth County, KS, Hydrologic Unit 10270104.

Equipment: An YSI 6600 water-quality monitor equipped with sensors for water temperature, specific conductance, dissolved oxygen, pH, turbidity, and chlorophyll was installed from August 2012 through May 2014. From June 2014 to the present (2015) a Xylem YSI EXO2 water-quality monitor equipped was installed with sensors for water temperature, specific conductance, dissolved oxygen, pH, turbidity, and chlorophyll. The monitor is housed in a 4-inch diameter galvanized steel pipe. Readings from the water-quality monitor are recorded every 15 minutes and transmits data by way of satellite, hourly.

Date model was created: October 15, 2015

Model calibration data period: July 19, 2012 – June 29, 2015

Model application date: July 19, 2012 onward

Model-Calibration Dataset

All data were collected using U.S. Geological Survey (USGS) protocols and are stored in the National Water Information System (NWIS) database. Linear regression models were developed using the open-source software package "R." Explanatory variables selected as inputs to linear regression were physicochemical properties: specific conductance, pH, water temperature, dissolved oxygen, turbidity, chlorophyll fluorescence, and streamflow. Seasonal components (sine and cosine variables) were also evaluated as explanatory variables in the models to determine if seasonal changes affected the model. All combinations of physicochemical properties and a seasonal component were evaluated to determine which combinations produced the best models.

The final selected regression model is based on 54 concurrent measurements of Alk concentration and specific conductance (SC) collected from July 19, 2012 through June 29, 2015. Samples were collected throughout the range of continuously observed hydrologic conditions. No samples were below laboratory detection limits. Summary statistics and the complete model-calibration dataset are provided below. Studentized residuals from the final model were inspected for values greater than 3 or less than negative 3. Values outside of that range are considered potential outliers and are investigated. One sample, June 11, 2014, was found to have potential errors in collection and processing, and has been removed from the dataset. All other potential outliers were not found to have errors associated with collection, processing, or analysis, and were therefore considered valid.

Alkalinity Sampling Details

Cross-section samples are typically collected either from the downstream side of the bridge or instream within 100 feet of the bridge. The equal-width-increment (EWI) method is used, and samples typically are composited for analysis. Cross-section samples are collected every 2 weeks from March through October, once a month from November through February, and during selected runoff events. A FISP US DH-95, D-95, or D-96A1 depth integrating sampler is used from the bridge; and a DH-81 or DH-95 hand sampler is used for boat samples. Samples are analyzed for Alk concentration at the USGS National Water Quality Laboratory in Lakewood, Colorado.

Model Development

Regression analysis was done using R by examining SC, streamflow, and other continuously measured data as explanatory variables for estimating Alk concentration. A variety of models that predict Alk, $(Alk)^2$, $\sqrt{Alk}$ and models that predict $\log_{10}(Alk)$ were evaluated. The distribution of residuals was examined for normality, and plots of residuals (the difference between the measured and computed values) as compared to computed Alk were examined for homoscedasticity (meaning that their departures from zero did not

change substantially over the range of computed values). This comparison lead to the conclusion that the most appropriate and reliable model would be one that estimated $\log_{10}$(Alk).

Alk was selected as the best predictor of Alk based on residual plots, relatively high adjusted coefficient of determination (adjusted R^2) and relatively low model standard percentage error (*MSPE*), prediction error sum of squares (PRESS), and Mallow's C_p. Values for all of the afore mentioned statistics and metrics were computed for various models and are included below along with all relevant sample data and more in-depth statistical information.

Model Summary

Summary of final regression analysis for Alk concentration at site number 06892350.

Alk concentration-based model:

$$\log_{10}(Alk) = 0.567 \times \log_{10}(SC) + 0.616$$

where

Alk = Alkalinity in milligrams per liter (mg/L) as calcium carbonate ($CaCO_3$); and,

SC = specific conductance in microsiemens per centimeter at 25 degrees Celsius (µs/cm)

SC makes physical and statistical sense as explanatory variables for Alk.

The log-transformed model may be retransformed to the original units so that Alk can be calculated directly. The retransformation introduces a bias in the calculated constituent. This bias may be corrected using Duan's Bias Correction Factor (BCF). For this model, the calculated BCF is 1.00. The retransformed model, accounting for BCF is:

$$Alk = SC^{0.567} + 4.13$$

Previous Models

No previous models have been published.

Alkalinity Concentration Record

The Alk record is computed using this regression model and stored at the National Real-Time Water Quality (NRTWQ) Web site. Data are computed at 15-minute intervals. The complete water-quality record can be found at http://nrtwq.usgs.gov/ks.

Remarks

None

R Output for Alkalinity; 06892350; Kansas River at De Soto, KS

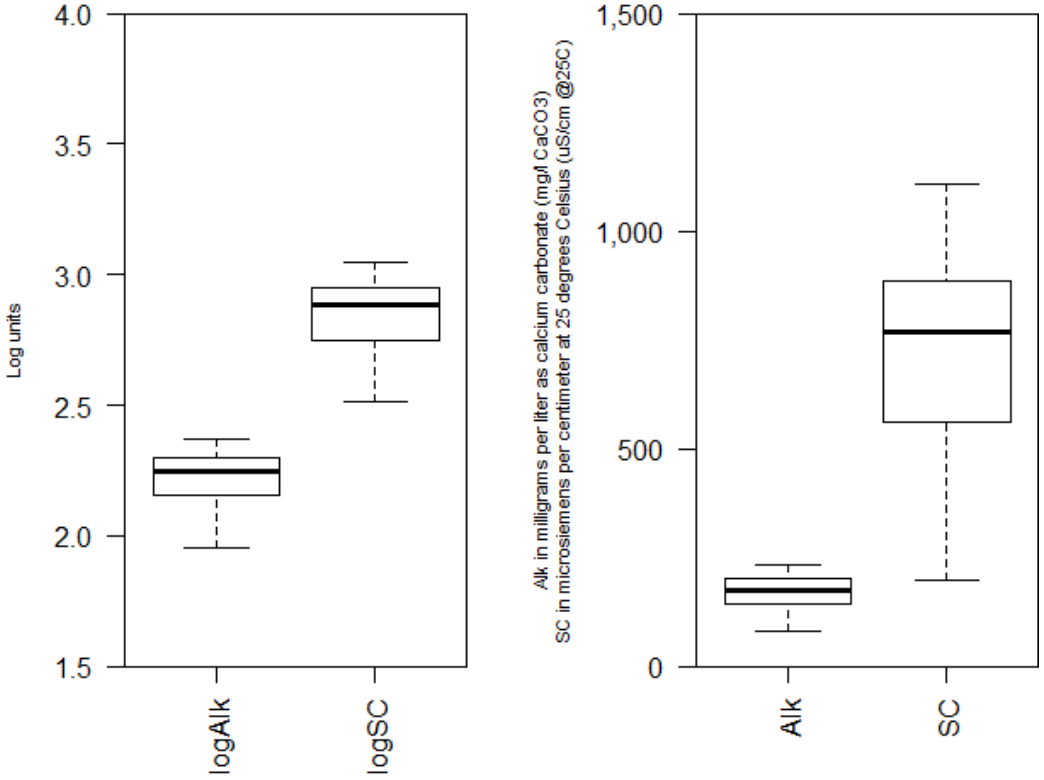
Model Form

logAlk = + 0.567 * logSC + 0.616

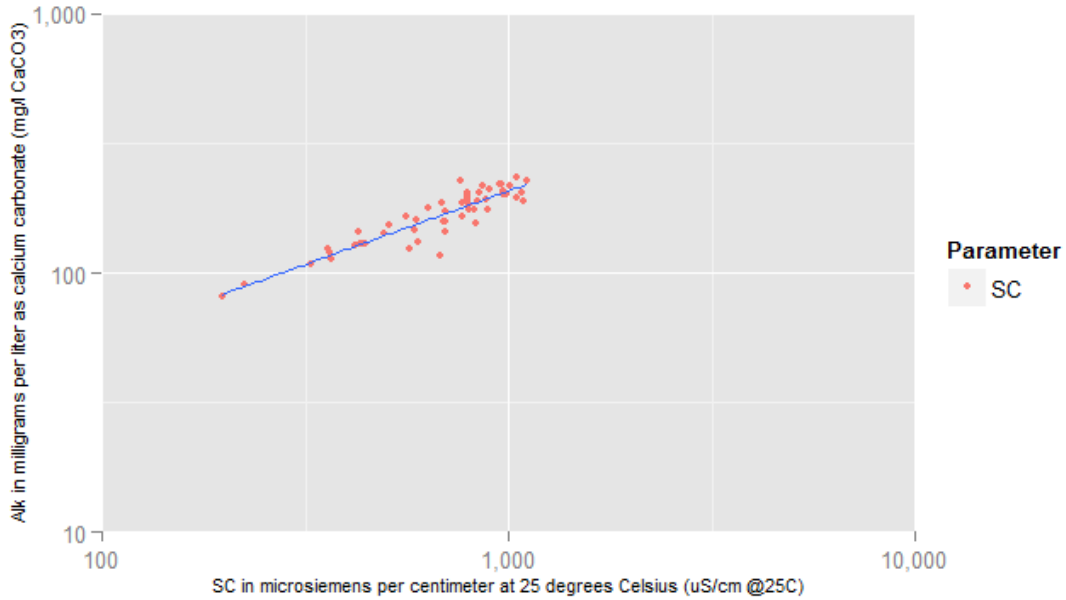
Variable Summary Statistics

	logAlk	logSC	Alk	SC
Minimum	1.91	2.29	80.4	197
1st Quartile	2.15	2.75	142.0	560
Median	2.24	2.89	175.0	769
Mean	2.22	2.83	170.0	720
3rd Quartile	2.30	2.95	200.0	886
Maximum	2.37	3.05	234.0	1110

Box Plot(s) of sample data



Exploratory Plot



Model Calibration

Basic Data

Number of Observations	54
Standard error (RMSE)	0.0434
Upper Model standard percentage error (MSPE)	10.5
Lower Model standard percentage error (MSPE)	9.52
Coefficient of determination (R^2)	0.839
Adjusted Coefficient of Determination (Adj. R^2)	0.836
Bias Correction Factor (BCF)	1

Explanatory Variables

	Coefficients	Standard Error	t value	Pr(>|t|)
(Intercept)	0.616	0.0976	6.32	6.06e-08
logSC	0.567	0.0344	16.50	2.88e-22

Correlation Matrix

	Intercept	E.vars
Intercept	1.000	-0.998
E.vars	-0.998	1.000

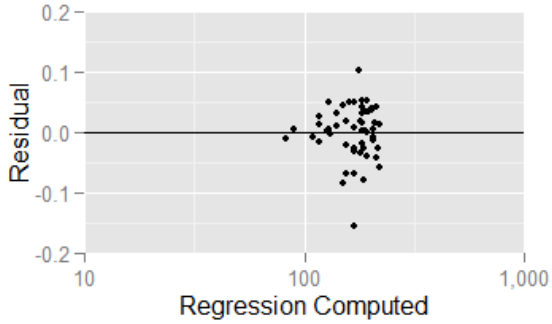
Test Criteria

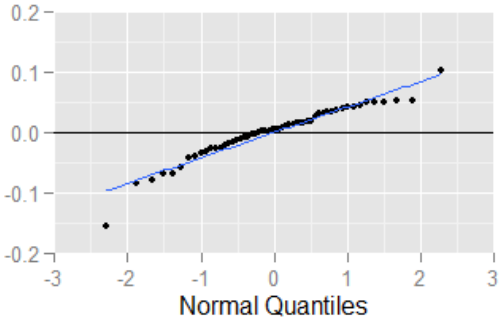
Leverage	Cook's D	DFFITS
0.05555556	0.10557024	0.27216553

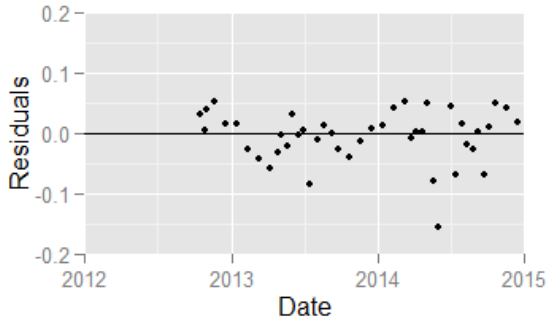
Flagged Observations

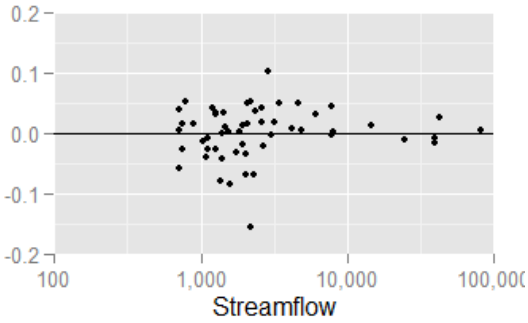
	logAlk	Estimate	Residual	Standard Residual	Studentized Residual	Leverage	Cook's D	DFFITS
4/8/2013 11:00	2.279	2.338	-0.058590	-1.38100	-1.39400	0.04604	0.0460300	-0.30610
7/15/2013 10:40	2.094	2.178	-0.083980	-1.95500	-2.01200	0.02185	0.0426900	-0.30060
8/5/2013 11:30	1.906	1.917	-0.011010	-0.28290	-0.28040	0.19750	0.0098510	-0.13910
8/19/2013 10:00	2.079	2.068	0.011320	0.26930	0.26690	0.06343	0.0024560	0.06946
5/19/2014 12:20	2.191	2.271	-0.079960	-1.86300	-1.91000	0.02373	0.0421900	-0.29780
6/2/2014 13:00	2.067	2.222	-0.155600	-3.61600	-4.13900	0.01854	0.1235000	-0.56890
1/12/2015 11:30	2.353	2.250	0.102800	2.39100	2.51000	0.02037	0.0594500	0.36200
5/18/2015 15:30	2.032	2.040	-0.008007	-0.19230	-0.19060	0.08134	0.0016380	-0.05670
6/6/2015 19:50	1.952	1.949	0.003761	0.09456	0.09365	0.16150	0.0008611	0.04110
6/15/2015 14:50	2.090	2.064	0.025670	0.61150	0.60780	0.06539	0.0130800	0.16080
6/29/2015 12:30	2.052	2.070	-0.017660	-0.41980	-0.41650	0.06221	0.0058460	-0.10730

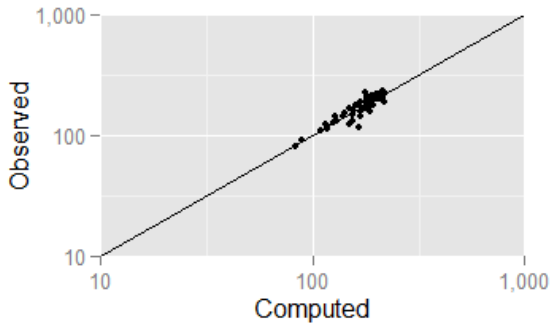
Statistical Plots

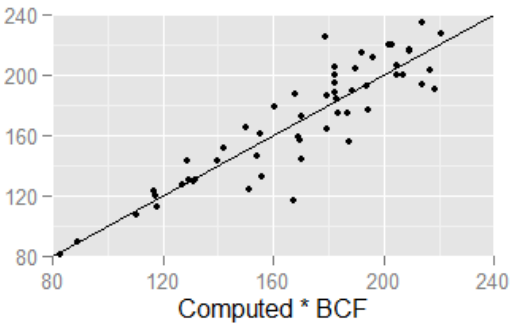


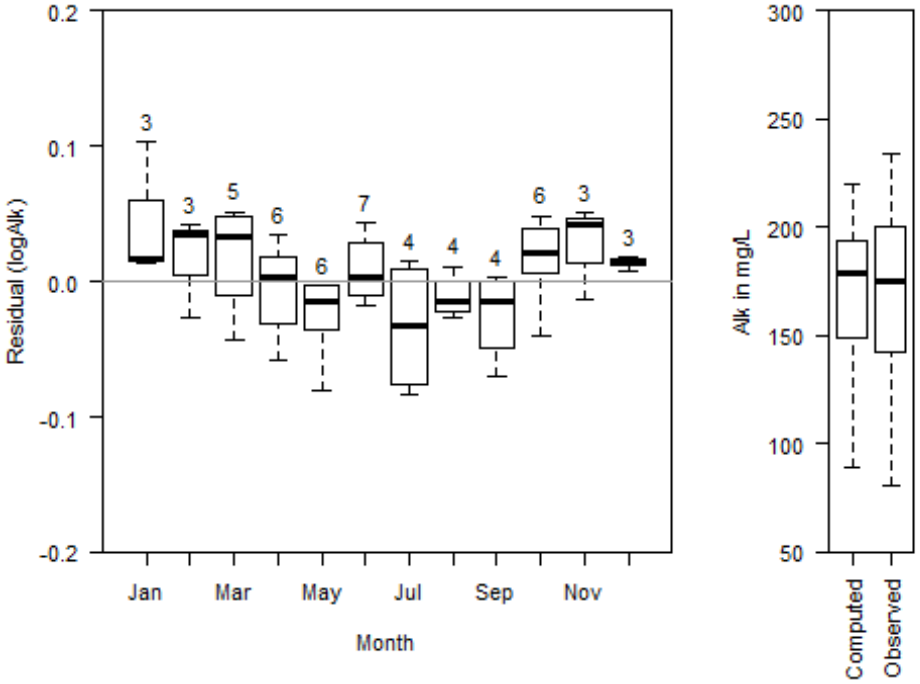


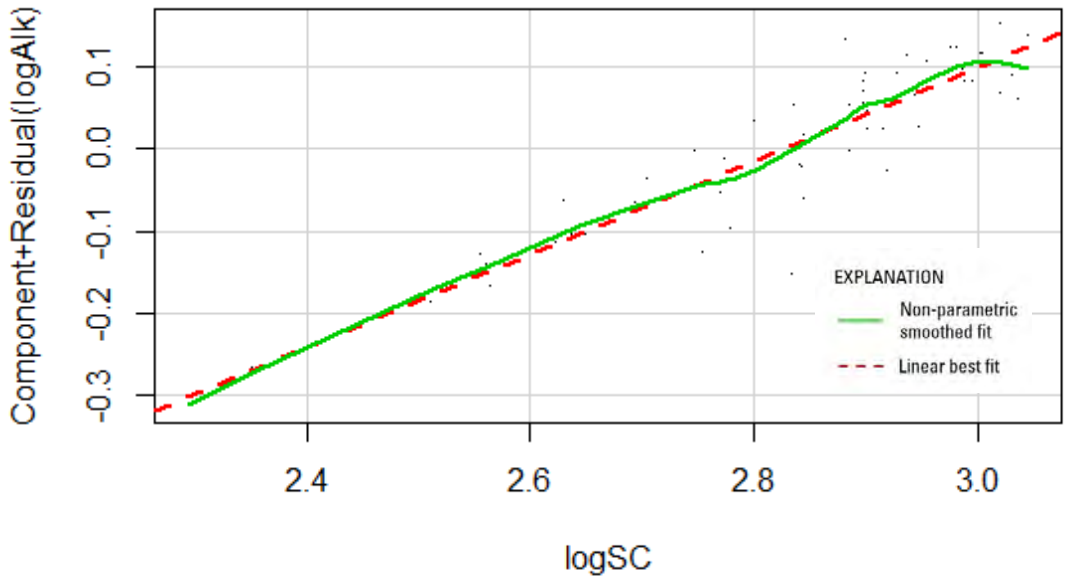









Residual (logAlk)
0.2
0.1
0.0
-0.1
-0.2
3
3
5
6
6
7
4
4
4
6
3
3
Jan
Mar
May
Jul
Sep
Nov
Month
Alk in mg/L
300
250
200
150
100
50
Computed
Observed


Component+Residual(logAlk)
0.1
0.0
-0.1
-0.2
-0.3
2.4
2.6
2.8
3.0
logSC
EXPLANATION
Non-parametric smoothed fit
Linear best fit

Models considered

Model Formula	Number of Variables	Standard Error	R2	Adjusted R2	Cp	PRESS	VIF	MSPE
logAlk ~ logSC	1	0.04343	83.9	83.59	23.04	0.1031	1	± 10
logAlk ~ SC	1	0.04853	79.89	79.5	41.2	0.133	1	± 11
logAlk ~ logTurb	1	0.05067	78.08	77.66	49.42	0.1436	1	± 12
logAlk ~ logSC + cos2DY	2	0.03756	88.19	87.72	5.576	0.08059	1.583	± 8.7
logAlk ~ Temp + logSC	2	0.03848	87.6	87.11	8.24	0.08459	1.401	± 8.9
logAlk ~ logSC + logTurb	2	0.0412	85.79	85.23	16.47	0.09559	5.212	± 9.5
logAlk ~ logQ + logSC + cos2DY	3	0.03692	88.81	88.14	4.745	0.07815	4.364	± 8.5
logAlk ~ SC + logSC + cos2DY	3	0.03728	88.59	87.91	5.74	0.07972	22.19	± 8.6
logAlk ~ logSC + logTurb + cos2DY	3	0.03747	88.47	87.78	6.284	0.08184	5.23	± 8.6
logAlk ~ logQ + SC + logSC + cos2DY	4	0.03609	89.52	88.67	3.512	0.07599	4.617	± 8.3
logAlk ~ logQ + logSC + Turb + cos2DY	4	0.03646	89.31	88.43	4.504	0.07666	4.769	± 8.4
logAlk ~ logQ + logSC + logTurb + cos2DY	4	0.03655	89.25	88.37	4.746	0.07813	4.471	± 8.4
logAlk ~ logQ + SC + logSC + logTurb + cos2DY	5	0.0359	89.84	88.79	4.056	0.07558	4.689	± 8.3
logAlk ~ Q + logQ + logSC + logTurb + cos2DY	5	0.03622	89.66	88.58	4.896	0.07881	3.6	± 8.4
logAlk ~ Q + logQ + SC + logSC + cos2DY	5	0.03634	89.59	88.51	5.209	0.07737	4.909	± 8.4
logAlk ~ logQ + logTemp + SC + logSC + logTurb + cos2DY	6	0.03607	89.96	88.68	5.543	0.07748	5.376	± 8.3
logAlk ~ Q + logQ + SC + logSC + logTurb + cos2DY	6	0.03614	89.92	88.64	5.698	0.07736	4.917	± 8.3
logAlk ~ logQ + Temp + SC + logSC + logTurb + cos2DY	6	0.03618	89.9	88.61	5.808	0.07693	4.94	± 8.3

Data

	Date	logAlk	logSC	Alk	SC	Computed logAlk	Computed Alk	Residual	Normal Quantiles
1	2012-10-15	2.289	2.898	194.7	791	2.259	182.3	0.0307	0.615
2	2012-10-26	2.315	2.987	206.4	971	2.309	204.7	0.00559	0.0694
3	2012-10-29	2.342	2.977	219.6	948	2.303	202	0.0384	0.926
4	2012-11-19	2.332	2.938	214.9	867	2.281	192	0.051	1.66
5	2012-12-17	2.335	3.004	216.1	1010	2.319	209.3	0.0158	0.354
6	2013-01-14	2.334	3.003	215.9	1008	2.318	209.1	0.016	0.404
7	2013-02-11	2.308	3.031	203.1	1075	2.334	216.9	-0.0264	-0.731
8	2013-03-11	2.286	3.021	193.1	1050	2.328	214	-0.0426	-1.17
9	2013-04-08	2.279	3.037	190.1	1090	2.338	218.6	-0.0586	-1.26
10	2013-04-25	2.196	2.843	157.1	697	2.228	169.6	-0.0313	-0.926
11	2013-05-06	2.114	2.649	130.1	446	2.118	131.7	-0.00336	-0.305
12	2013-05-20	2.164	2.77	145.9	589.2	2.186	154.2	-0.0221	-0.672
13	2013-06-03	2.181	2.706	151.8	508	2.15	141.8	0.0316	0.672
14	2013-06-17	2.112	2.644	129.5	441	2.115	130.9	-0.00259	-0.257
15	2013-07-01	2.115	2.637	130.2	433.5	2.111	129.6	0.00397	0.0231
16	2013-07-15	2.094	2.755	124.1	569.4	2.178	151.3	-0.084	-1.89
17	2013-08-05	1.906	2.294	80.45	197	1.917	82.9	-0.011	-0.455
18	2013-08-19	2.079	2.561	119.9	363.8	2.068	117.4	0.0113	0.21
19	2013-09-09	2.285	2.945	192.6	882	2.285	193.9	-0.000801	-0.21
20	2013-09-23	2.242	2.917	174.6	826	2.269	186.8	-0.0273	-0.858
21	2013-10-21	2.246	2.947	176.2	886	2.287	194.4	-0.0406	-1.08
22	2013-11-18	2.301	2.996	200	990.6	2.314	207	-0.013	-0.506
23	2013-12-16	2.236	2.846	172.3	701	2.229	170.2	0.00736	0.116
24	2014-01-13	2.355	3.045	226.7	1110	2.342	220.8	0.0134	0.257
25	2014-02-10	2.37	3.021	234.3	1050	2.328	214	0.0414	1
26	2014-03-10	2.311	2.899	204.6	793.3	2.259	182.6	0.0515	1.89
27	2014-03-24	2.3	2.988	199.5	973.2	2.31	205	-0.00974	-0.404
28	2014-04-07	2.276	2.925	188.9	841.2	2.274	188.7	0.00242	-0.163
29	2014-04-21	2.264	2.902	183.7	797.8	2.261	183.1	0.00334	-0.0694
30	2014-05-05	2.272	2.836	187.2	684.9	2.223	168	0.0491	1.5
31	2014-05-19	2.191	2.919	155.1	830.2	2.271	187.3	-0.08	-1.66
32	2014-06-02	2.067	2.834	116.6	682.3	2.222	167.6	-0.156	-2.29

33	2014-06-30	2.217	2.748	164.8	559.7	2.174	149.8	0.0434	1.17
34	2014-07-14	2.159	2.845	144.2	699.3	2.228	170	-0.0694	-1.37
35	2014-07-28	2.274	2.899	187.8	792.8	2.259	182.5	0.0145	0.305
36	2014-08-11	2.242	2.902	174.7	798.2	2.261	183.2	-0.0186	-0.615
37	2014-08-25	2.2	2.841	158.4	693.4	2.226	169.2	-0.0265	-0.793
38	2014-09-08	2.106	2.623	127.5	420.2	2.103	127.4	0.00253	-0.116
39	2014-09-22	2.122	2.779	132.3	601.5	2.191	156.1	-0.0697	-1.5
40	2014-10-06	2.154	2.695	142.5	495.5	2.144	139.8	0.0103	0.163
41	2014-10-20	2.156	2.631	143.2	427.8	2.107	128.7	0.0486	1.37
42	2014-11-17	2.301	2.899	199.9	792.8	2.259	182.5	0.0416	1.08
43	2014-12-15	2.207	2.775	161	595.1	2.189	155.1	0.0182	0.455
44	2015-01-12	2.353	2.883	225.2	763	2.25	178.6	0.103	2.29
45	2015-02-09	2.341	2.981	219.2	958	2.306	203.2	0.035	0.858
46	2015-03-09	2.325	2.955	211.2	902	2.291	196.3	0.0337	0.731
47	2015-03-23	2.252	2.801	178.7	633	2.204	160.6	0.0483	1.26
48	2015-04-06	2.31	2.928	204.3	848	2.276	189.6	0.0345	0.793
49	2015-04-20	2.271	2.886	186.5	769.5	2.252	179.4	0.0188	0.506
50	2015-05-04	2.216	2.886	164.4	769.3	2.252	179.4	-0.0359	-1
51	2015-05-18	2.032	2.512	107.6	325.1	2.04	110.1	-0.00801	-0.354
52	2015-06-07	1.952	2.351	89.63	224.5	1.949	89.27	0.00376	-0.0231
53	2015-06-15	2.09	2.555	123	359	2.064	116.5	0.0257	0.56
54	2015-06-29	2.052	2.565	112.7	366.9	2.07	117.9	-0.0177	-0.56

Definitions and National Water Information System (parameter code)

Alk: Alkalinity in mg/L CaCO3 (39087)
SC: Specific conductance in uS/cm @25C (00095)